THE INTERDISCIPLINARY JOURNAL OF PROBLEM-BASED LEARNING

Conducting Problem-Based Learning Meta-Analysis: Complexities, Implications, and Best Practices

Andrew Walker (Utah State University)
Heather Leary (Brigham Young University)

IJPBL is Published in Open Access Format through the Generous Support of the School of Education at Indiana University, the Jeannine Rainbolt College of Education at the University of Oklahoma, and the Center for Research on Learning and Technology at Indiana University.

The Interdisciplinary Journal of Problem-based Learning

2023 SPECIAL ISSUE

Conducting Problem-Based Learning Meta-Analysis: Complexities, Implications, and Best Practices

Andrew Walker(Utah State University)
Heather Leary (Brigham Young University)

ABSTRACT

Almost fifty years ago, Barrows (1986) claimed that problem-based learning (PBL) was broad enough that a single methodological description was not possible. It has only grown more complex since. In addition to meaningful variations of PBL, there are several related problem-centered pedagogies, such as case-based learning, project-based learning, and inquiry-based learning, among others. Even within PBL, primary research is conducted using a wide variation of measurement approaches, with diverse audiences, in a myriad of disciplines. The inherent complexity of PBL research can present some unique challenges to meta-analysis, such as multiple definitions of "control groups." PBL research also intersects with common meta-analysis challenges such as preexperimental, and multiple treatment designs. This article will explore best practices for conducting meta-analysis using a modest expansion of data and new analyses based on Leary et al. (2013). Readers will see major sections of a meta-analysis alongside a running worked example, enabling a discussion of meta-analytic methods unique to a PBL context.

Keywords: problem-based learning, meta-analysis, best practices, challenges

Problem-based learning (PBL) is an educational approach in which students learn by working on real-world problems. As readers are no doubt aware, it has broadened into many disciplines since its origins in medical education (Barrows, 1996) and has spawned enough meaningful variations that one of the early proponents of PBL, Howard Barrows, developed a taxonomy with which to classify them (Barrows, 1986). A more modern perspective would discuss these variations as problem-centered pedagogies and expand to include additional approaches, such as inquiry-based learning (Hmelo-Silver et al., 2007). PBL has matured and evolved over a period of decades with a large corpus of primary research informed by both scholars and practitioners. As might be expected, several scholars have undertaken meta-analyses of this literature.

Meta-analysis shares a decades-long history with PBL (Glass, 1976) and has also grown and matured over the years (Glass, 2006). It has garnered enough popularity to go beyond statistical software like R, which pride themselves on user-created packages and extensions to be incorporated into software as ubiquitous as SPSS. There is even a specialized Comprehensive Meta-Analysis application. As an aside, R is recommended as it is often used by methodologists to create packages and pilot new analyses techniques. The tool is thus up to date and also free. Leary and Walker (2018) is a light introduction to how meta-analysis has evolved over time to include metasynthesis, qualitative metasynthesis, network, and Bayesian network, in addition to traditional meta-analysis. There are also several excellent books, including the following, for readers looking to get additional details. Lipsey and Wilson (2001) offer a good introduction and reference to key topics, companion data, and sample analyses

https://doi.org/10.14434/ijpbl.v17i2.35844

and valuable details on effect size calculation. Cooper et al. (2019) have a handbook in its third edition with broad coverage and details for individual topics, from framing a problem appropriate for meta-analysis and searching to reporting recommendations. Borenstein et al. (2021) are in their second edition of a good text for a meta-analysis class, with robust coverage and few a priori assumptions about familiarity with meta-analysis.

While there is a large volume of PBL specific meta-analyses and a much larger set of methods papers, innovations, handbooks, and textbooks for meta-analysis, generally there is not much coverage of meta-analytic methods specific to the context of PBL. The purpose of this article is to provide guidance on meta-analysis using examples from the PBL literature as well as to highlight needs for future work.

Types of Reviews

Any good review work should have early coverage of existing reviews (Boote & Beile, 2005). Interrogating the quality, coverage, and methods used can then inform new review work. The goal of research reviews, sometimes known as research synthesis, is to build a bigger picture on a topic from individual studies. Reviews provide information on the previous research conducted and where there are gaps and future needs (Cooper et al., 2019). Reviews can be quantitative (meta-analysis), qualitative (metaethnography and qualitative metasynthesis), and mixed-methods (metasynthesis; Leary & Walker, 2018). Much like primary research, replication work is welcome and needed in reviews. An update to a review that is over ten years old can be a valuable contribution, especially given the way the PBL landscape changes over time.

In the case of PBL, the existing review space is extensive. Among the earliest efforts, Albanese and Mitchell (1993) stopped short of conducting a meta-analysis and completed a systematic review of medical education PBL articles. This allowed them to report on research questions addressed in only a handful of studies, such as the relative costs of PBL. A systematic review or a combination with meta-analysis may still be appropriate for PBL reviewers with rigid inclusion criteria. Note that systematic reviews are often confused for meta-analyses, and some use the terms synonymously. Meta-analyses were conducted in medical education around this time. Vernon and Blake (1993) did a hybrid systematic review and meta-analysis, with a systematic review of five studies on evaluative (e.g., attitudes, opinions, or attendance) and process outcomes in addition to meta-analyses of knowledge tests and clinical performance. Kalaian et al. (1999) conducted an advanced form of meta-regression of mean differences in medical education by having a tight focus on a common set of outcomes, the standardized National Board of Medical Examiners I and II. In addition to exploring predictive relationships, their Hierarchical Linear Modeling (HLM) approach adjusts for the common study of origin when there are multiple outcomes. Subsequent efforts went beyond medical education to literature in other disciplines (Dochy et al., 2003), including another combination systematic and meta-analytic review (Newman, 2003). This work paved the way for a reframing of this complicated measurement space with a meta-analysis by Gijbels et al. (2005) that incorporated the work of Sugrue (1993), who conceptualized an assessment framework for problem solving. In similar fashion, Walker and Leary (2009) examined the implementation and used Barrow's taxonomy (1986) to explore the posited gains of PBL. From here, the review space matures to include a metasynthesis of meta-analyses, a far more robust and well-visualized version of this coverage of the review space (Strobel & Barneveld, 2009), as well as more niche looks at the impact of tutors (Leary et al., 2013). PBL has progressed enough that it may be part of meta-analysis efforts in related literature, such as all problem-centered pedagogies paired with computer-based scaffolding (Belland et al., 2017a; Belland et al., 2017b). It should be noted that this coverage of existing meta-analytic reviews is far from complete—it introduces one standalone systematic review largely to help readers differentiate between systematic and meta-analytic reviews. Depending on the scientific argument being made, a good review of related literature may include systematic, meta-analyses, and narrative reviews (e.g., Berkson, 1993). Finally, quality of the review work is not addressed as recommended, to conserve article length.

Sections of a Good Meta-Analysis

While there are times when variation is important, most strong meta-analyses include some specific sections of the paper. What follows are details on these common sections. They are drawn from Leary et al. (2013), a meta-analysis of tutor experience and background in PBL. Use of a single reference allows readers to have a common context for each section. Of note, some of the sections consist of new or re-analyses of these data to adopt best practices. There is also change in the data used. Leary et al. (2013) excluded studies that used automated, as opposed to human, tutors and excluded some studies for incomplete data across all coding. Focusing analyses on a handful of coding categories increased the total studies and outcomes.

Introduction

Much like the one modeled in this and other articles in the special issue, a good introduction should walk readers through the logic of prior work and clearly articulate the contribution of a new meta-analysis. A strong meta-analysis may investigate a precise portion of the PBL literature, such as medical education (Kalaian et al., 1999; Vernon & Blake, 1993). It is important to be precise with key concepts and how they are operationalized, most often as a form of measurement (Cooper et al., 2019). This was the methodological leap forward contributed by Gijbels et al. (2005), which categorized assessment outcomes using a simplified version of Sugrue's (1995) problem solving framework. Subsequent PBL reviews, especially those looking broadly across disciplines, are likely to run into a similarly diverse set of measurement outcomes, and it would be wise to adopt a measurement framework. Coding for problem types from Jonassen (2000) or PBL implementations from Barrows (1986) is another example of operationalizing concepts important to the literature, as used by Walker and Leary (2009). All these approaches provide a clear sense for the contribution of the meta-analysis and also aid with subsequent searching, inclusion/exclusion, and coding phases. The PBL literature is mature enough that it has shifted from medical education specific meta-analyses (Kalaian et al., 1999; Vernon & Blake, 1993) to reviews that cast a wide disciplinary net (Dochy et al., 2003). Going forward, there may be a benefit to more niche reviews, whether that means a focus on specific learner populations, subject areas, or measurement instruments. In terms of the example review, Leary et al. (2013) represents the first attempt to conduct a meta-analysis of the training and subject matter expertise of PBL tutors; due to a lack of available studies, the only prior work in this space (Albanese & Mitchell, 1993) is a brief narrative review as part of a larger review effort.

Research Questions

As with any research study, the research questions drive the research activities. Research questions for meta-analysis should aim to ask about summarized evidence on a topic from multiple related studies. Meta-analysis research questions target understanding, through quantitative assessment, the value of different intervention features (Hansen et al., 2022). Scoping a meta-analysis is important, and the research questions should reflect a targeted topic and make a clear contribution. Too few articles might limit the impact of the meta-analysis technique, while too many articles can cause management issues. Researchers must balance the relevance of articles aligned with the meta-analysis research questions and inclusion/exclusion criteria for a manageable volume of primary research studies.

Historically in PBL, research questions where meta-analysis were used have been focused on comparisons of PBL treatments versus a control group (Walker & Leary, 2009) or a specific topic related to PBL, such as tutors in PBL (Leary et al., 2013). The PBL literature and research community benefit greatly from meta-analysis contributions. These reviews provide cumulative reporting on many areas of PBL in the literature. Generating research questions is often an iterative process in relation to a priori and emergent coding decisions over the course of the meta-analysis. With the exception of a Bayesian network meta-analysis, most meta-analyses will have a general effectiveness research question addressed by a summary effect size. From there, depending on the meta-analysis, there may be group difference questions or predictive questions centered around meta-regression. Both borrowing from and expanding on the example study Leary et al. (2013), research questions could include:

1. "What is the overall contribution of PBL across the entire range of PBL tutor training, experience, and expertise?" to explore a summary effect;
2. "How do variations in expertise (novice, expert, mixed, automated) impact student learning?" to explore subgroup differences; or
3. "Which tutor expertise categories (novice, expert, mixed, automated) predict student learning?" to use meta-regression to investigate study quality.

While there may and should be important analyses that are not tied to a specific research question, it is recommended that primary analyses be tied to specific research questions.

Methods

Meta-analysis methods should be detailed enough to support replication, just like any primary research study. Below are five steps to use when conducting a meta-analysis. As the PBL literature base continues to grow and meta-analysis techniques are updated and refined, it is vital to follow these steps carefully.

Inclusion and Exclusion Criteria

In meta-analytic research, it is recommended for researchers to include all relevant studies (Cooper et al., 2019). In many ways, a meta-analysis is intended to be a full population study where the subjects are all relevant primary research. Including all relevant studies reduces bias and provides coverage of the topic in question, so the most cited studies are not the ones that get the most attention.

Good coverage, which is helped through the search process, allows for gray literature to be included and minimizes the "file drawer problem" (Rosenthal, 1979). Gray literature can include technical reports, dissertations and thesis, or conference papers that never became journal publications. Often, these manuscripts are left in the metaphorical file drawer because they contain nonsignificant or unexpected findings. In a more subtle variation, scholars may omit results within a published study (such as only reporting statistically significant results). Researchers must decide what variables and study characteristics are important to include in the meta-analysis as well as what manuscripts to exclude. Both are important to judge whether a manuscript should be part of the final analysis, and both inclusion and exclusion criteria should be clearly reported.

The specific inclusion criteria will depend on the research questions and scoping of the meta-analysis. Generally, meta-analyses include published literature, unpublished literature, manuscripts with quantitative data and analysis, or a control versus a treatment. More specifically for PBL, inclusion criteria might consist of a specific grade level (e.g., K–12 or higher education), discipline (e.g., medical education, nursing, science, or history), how PBL is defined, or if a scaffold is used or not. Just as important is the exclusion criteria. These depend on the research questions and focus of the meta-analysis, but might include qualitative studies, studies that fail to report enough information to calculate an effect size, or omission of a PBL feature such as tutors. Justification for inclusion/exclusion decisions should be detailed, keeping the meta-analysis process transparent. Authors should be systematic and document all activities, especially as inclusion and exclusion intersects with searching.

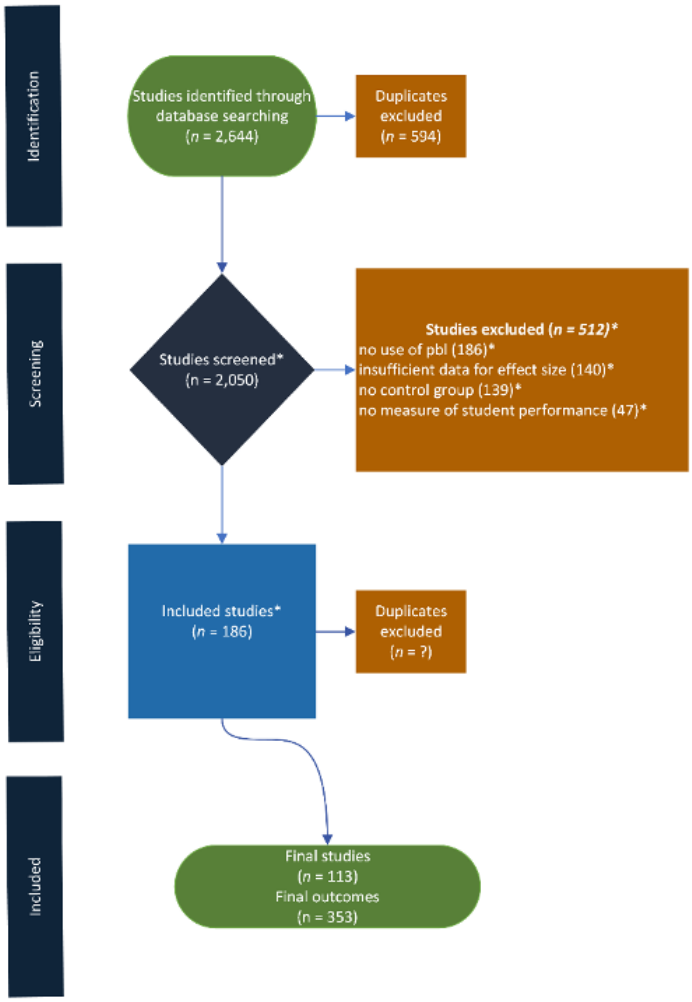
Search Process

The search process for manuscripts to include in a meta-analysis is quite similar to a systematic literature review. It should be transparent, replicable, and fully cover relevant research (Gusenbauer & Haddaway, 2020). The first strategy in the search process should be identifying previous literature reviews and meta-analyses on the same or similar topic. These are rich information sources and can help a researcher identify relevant manuscripts as a starting point. The next step is keyword searches in relevant databases (Glanville, 2019). Gusenbauer and Haddaway (2020) offer a strong empirical assessment of various databases specifically for the purpose of systematic review and offer strong opinions to avoid Google Scholar as replicating searches at a subsequent time point is not possible. PBL meta-analysts should consider ERIC, PsychInfo, Digital Dissertations, ScienceDirect, Scopus, and Web of Science, and those interested in medical education should add Nursing & Allied Health, Public Health Database from ProQuest, Virtual Health Library, and PubMed. Keyword searches might also capture gray literature, which are important to include in meta-analyses. To reduce bias in the search process, researchers should use multiple databases and, when possible, trained librarians who are helpful in identifying relevant strategies and databases to use in the search.

With PBL extending well beyond just the medical literature, identifying the best databases to search is crucial for staying within scope of the meta-analysis while also identifying all relevant manuscripts. Using technological abilities, such as semi-automated searches or forward citations (sometimes seen as "cited by") can be helpful. There are also specific journals (such as the Interdisciplinary Journal of Problem-Based Learning), leading scholars' websites (e.g., Cindy Hmelo-Silver's faculty profile from Indiana University, https://education.indiana.edu/about/directory/profiles/hmelo-silver-cindy.html), or Google Scholar pages (e.g., Sofie M. M. Loyens' Google Scholar profile, https://scholar.google.nl/citations?user=NqC7qlkAAAAJ&hl=en) that list manuscripts that might be relevant to the topic being explored through meta-analysis.

The search path for Leary et al. (2013) is a complicated one, building forward from Walker & Leary (2009) which included a primary research author survey for referrals in addition to subsequent searching. A scoping review was conducted, which is a good practice for estimating the scope of work for a proposed meta-analysis. For the scoping review, PsychInfo, ERIC, Education Abstracts, Education Full Text, and Medline were searched using the terms "problem based learning" OR "problem-based learning" AND "control". Any limiters for research reports or empirical research were utilized. Results were restricted to 2013 and earlier. After duplicate removals of 2,644 search returns by Ebscohost, a total of 2,050 references were uploaded to rayyan (see http://rayyan.ai) for sampling. The set was sorted by year so that relevance was not a factor, and then every 50 articles (n = 44) were pre-pass screened, meaning an examination of only titles and abstracts. The resulting promising articles from pre-pass (n = 15) were examined more closely for inclusion/exclusion criteria and to determine the number of acceptable studies (n = 4) and outcomes (n = 6) pulled from this small sample (n = 44) of the search returns. Keeping in mind this represents only 2% of the 2,050 search returns, the scoping review suggested 186 studies and 280 outcomes for final inclusion. A scoping review allows researchers to adjust their search and inclusion/exclusion criteria.

Since details about inclusion/exclusion decisions for Leary et al. (2013) are not available, the scoping review is also used to estimate the PRISMA diagram (see Figure 1). The scoping review, which did not identify any duplicate results among



Note. *Estimated counts from a scoping review of 2% of search returns

Figure 1. PRISMA Diagram as Estimated from a Scoping Review

the four studies found for likely inclusion, comes close to approximating the final studies and outcomes. As is common in PBL, most of the estimated exclusions were due to the use of a problem-centered pedagogy that is not PBL (n = 186). Underreporting of results that prohibited effect size calculation (n = 140) was another major reason for exclusion, as is common with all meta-analyses. PRISMA diagrams in final meta-analyses should report carefully tracked numbers, as opposed to estimates.

Effect Size Calculations

For a detailed overview of various effect size calculation approaches, readers should examine Appendix B from Lipsey and Wilson (2001) as well as the companion website for actual calculations housed by the Campbell Collaboration (Wilson, n.d.). The book and website cover mean differences (Cohen's d) as well as correlation based (r) effect sizes of all forms, and the data needed to calculate them. Most meta-analyses that feature PBL examine mean differences between groups, especially between a control and a treatment group expressed as Cohen's d. There is a meta-analytic review that covers mean differences within groups (pre-post gains) for problem-centered pedagogies broadly, including PBL in the context of computer-based scaffolding (Belland et al., 2017a), but no efforts focusing specifically on gain scores with a focus on PBL. Also missing is an exploration of correlations within the PBL literature that could be especially helpful for affective or cognitive outcomes, since there are robust claims about the utility of gains in those spaces (Albanese, 2000) that deserve equally robust empirical analysis.

A discussion of magnitude of effect size is warranted. Going back to some early writings, Cohen (1988) expressed some understandable hesitancy in providing specific labels and ranges for effect sizes in an area of scholarship as broad as the social sciences. Cohen broadly outlined a range of small (d = 0.2), medium (d = 0.5), and large (d = 0.8) with some arguably problematic examples. For instance, Cohen cited the small advantage of men over women on a subtest of the Wechsler Adult Intelligence Scale. It is problematic to hold a binary view of gender, and an instrument that displays a gender bias should be interrogated. Nevertheless, it is possible to rely on Cohen's perspective that the social sciences are complex. Taking an intentional approach to measurement and research that considers diversity and equity as context for findings over universal and rigid thresholds is the best approach.

Specific to PBL research, magnitude has been discussed and debated. Colliver (2000) sets a high bar of d = 0.8–1.0 for a meaningful impact, citing that it is half of the two-sigma gain expected in one-on-one tutoring. Albanese (2000) counters with d = 0.5 as a reasonable expectation, citing it as a general midpoint for educational interventions as well as a threshold for pharmaceutical interventions. More likely, this will vary based on the PBL research of interest. Cohen's cautions of variability in social sciences are similar to the claims of Kraft (2020). Effect sizes will vary based on the reliability of measurement, the length of intervention, the age and potential maturation of learners, how soon after the intervention measurement is done, and if the study is small or done at scale with a general population. Kraft (2020) does suggest adjusted metrics of less than d = 0.05 as small, less than 0.20 as medium, and anything over 0.20 as large for preK12 populations. Those thresholds are quite low. Regardless of the thresholds used, context is an impactful consideration that should be supplied by any meta-analyst.

Coding

Coding in meta-analysis must be systematic and done with a designated coding sheet. This process, extracting information from manuscripts, takes up the bulk of the time to conduct meta-analyses. There is not a template or universal coding sheet that should be used because each meta-analysis has a different focus based on the research questions and meta-analytic method chosen. PBL researchers can review previous meta-analyses to gain an understanding of what these studies have used in the past for their coding sheet. At a minimum, the coding sheet should include a column for the full citation of the manuscript being coded or another identifier. After that, there are any number of columns to include to capture the extracted data from the manuscripts coded. What should be included in the coding sheet for a meta-analysis focused on PBL can vary widely, and writing all the potential coding columns or elements to include here would be difficult and something would be missed. A researcher should include all the characteristics of the studies that are important to answering the research questions so the data can be compared across all the manuscripts included in the meta-analysis.

In previously conducted meta-analyses, Newman (2003) included lengthy appendices of the coding sheet used and criterion captured. Some of this included the country of study, the discipline, the academic level, the length of intervention, whether PBL is an addition or part of the whole curriculum, the backgrounds of the tutors, how reliability was conducted, and the PBL description. Other previous meta-analyses in PBL have included a summary of the studies included with some of the elements coded (Albanese & Mitchell, 1993; Walker & Leary, 2009). This is a recommended best practice so that readers can look for patterns and get a sense of the relationships between coded features and outcomes. Some meta-analytic researchers publish their data collection and companion coding sheet in online archives that can be accessed and reviewed.

For the example case (Leary et al., 2013), one primary coding structure is the content expertise of the tutor. Content novices represented educational peers, while content experts had attained at least one educational step beyond the research participants (such as a Master's student tutoring undergrads). Emergent codes were added for situations where there was at least one tutor from each content expertise category, studies where the tutor background was missing data, and studies that employed an automated tutor via computer.

Choosing Meta-Analytic Method

With research questions and effect sizes selected, the next choice is a meta-analytic method. Hansen et al. (2022) identified four types of meta-analytic methods: (a) traditional univariate meta-analysis, (b) meta-regression analysis, (c) meta-analytic structural equation modeling, and (d) qualitative meta-analysis. There are also techniques to handle quasi-experimental and preexperimental designs that focus on gain scores, such as Bayesian network meta-analyses (e.g., Belland et al., 2017a). This section focuses on traditional meta-analysis and meta-regression. Leary and Walker (2018) offers an introduction to the wider scope of meta-analyses, including qualitative metasynthesis. The example study (Leary et al., 2013) is a univariate meta-analysis focused on mean differences between a control group and a group receiving a PBL intervention as part of the same research study.

Traditional Univariate Meta-Analysis. This is the traditional method for meta-analysis, with the weighted mean effect sizes for relationships (correlation based) or mean differences of the intervention being explored (Borenstein et al., 2021). Results consist of a summary effect size and, in the case of mean differences, should be reported alongside measures of heterogeneity and variability. It is important for meta-analysts to be clear about their choice of a fixed or random-effects model; this should be an a priori decision based on the nature of the included studies.

The example study (Leary et al., 2013) does a good job of being explicit about the use of random effects for subgroup comparisons. It could be strengthened by clarifying random effects as the model employed for all analyses and providing a justification for the choice. In this case, the breadth of study disciplines, target populations, variations in PBL intervention, and measures employed make it unlikely that a single true effect size is present, warranting the use of a random effects model. By contrast, the focused nature of Kalaian et al. (1999) does lend itself to a fixed effects model. Again, this should be a theoretical and a priori decision. Once a summary effect is determined, an ANOVA style z test can be used to determine if the results are significantly different from zero, or when subgroups are compared, it can be used to determine if any of the pairwise differences are significantly divergent from each other (Borenstein et al., 2021).

Visualizing the subgroup outcomes (see Figure 2) alongside key information like the point estimate, confidence intervals, and number of outcomes is a good practice. This is especially true when those subgroups are juxtaposed with the overall outcome estimate (in bold and using a filled diamond). In this case, the number of outcomes is prohibitive, but with fewer data points, a forest plot for each subgroup can help convey the level of heterogeneity within each group.

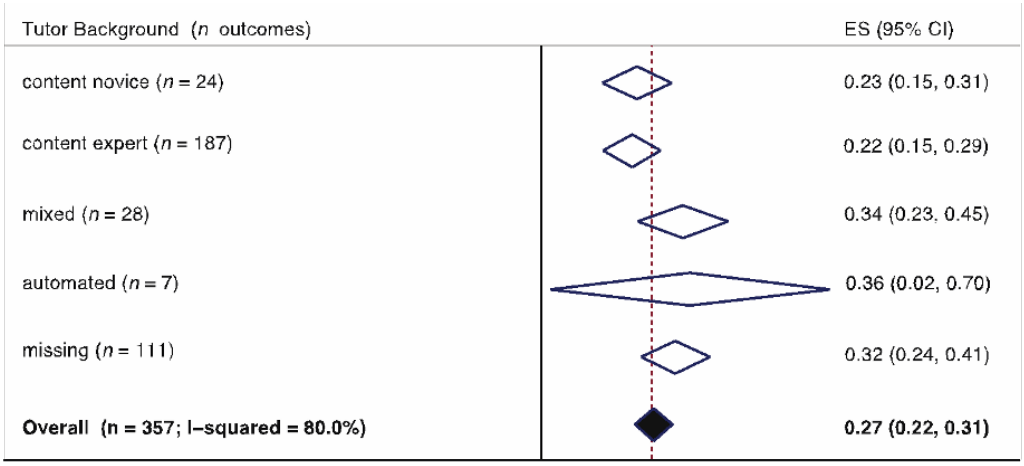


Figure 2. Subgroup Comparisons for Tutor Background

For the ANOVA style z test comparisons (Borenstein et al., 2021), even the largest mean difference between mixed (d = 0.34) and content experts (d = 0.22) failed to achieve statistical significance, z(213) = 1.80, p = 0.07. An appropriate conclusion is that tutor background alone is not able to explain variations in effect size outcomes; with this, there is not enough to support claims that recommend the use of expert or novice tutors.

Meta-Regression Analysis. Much like traditional multiple regression, meta-regression uses the effect size estimates as the outcome and coded variables as potential predictors. A final regression model is determined. Meta-analysts should be precise about the decision points of regression, such as the variable selection method and alpha values. The final regression model with R2 values is reported alongside significant predictor variables. Like with traditional meta-analysis, both a fixed and a random effects model is possible in meta-regression. Using the same tutor background data (Leary & Walker, 2013) in a dummy coded regression model with content novice held out as the reference group, no significant predictors of student learning were found, R2 = .01, Wald X2(4, 348) = 6.86, p = 0.14. This analysis used a random effects model with a restricted maximum likelihood estimator.

In contrast with typical linear regression, it is common to report Wald X2 as the model statistic rather than F; similarly, z-scores are reported rather than t-scores for predictors and the constant to reflect the fact that t-scores converge on z-scores with large sample sizes. Starting with content novices as the standard, only content experts have a lower estimated effect size—most of the explanatory power of the regression model is shared in common across all tutor backgrounds as expressed in the constant. Much like the ANOVA based subgroup comparison, there is no statistical significance, so variations in tutor background are not predictive of student performance.

Software

Most modern software packages contain at least basic tools for meta-analysis, including SPSS. For those who are willing to work in the command line, R is free and will also have the most up to date analysis techniques as methodologists start by writing new packages in the most accessible/

Tutor background	*b*	*se*	*z*	*p*	*ci.l*	*ci.u*
Content expert	-0.01	0.10	-0.11	0.91	-0.20	0.18
Mixed	0.16	0.13	1.32	0.19	-0.80	0.41
Automated	0.14	0.22	0.63	0.53	-0.29	0.56
Missing	0.11	0.10	1.09	0.27	-0.09	0.31
_cons	0.23	0.09	2.57	0.01	0.05	0.40

Table 1. Predictors of Student Learning

extensible platform. There is a niche software package titled "Comprehensive Meta-Analysis." While it positions itself as a one stop shop for all things meta-analysis, caution is warranted. Similar to SPSS, Comprehensive Meta-Analysis often presents results without walking through the need for intentional decision making.

Results

Much like the similarity across methods, there tend to be some common elements to meta-analysis results. The following sections walk through common result elements.

Heterogeneity

Several supporting analyses provide important context for meta-analytic results. The first is an assessment of heterogeneity. Beginning with Q, meta-analysts need to test a null hypothesis that all the observed studies share a common effect size. Importantly, this is a standardized measure that does not make claims about the magnitude of any variation between studies, so it should be used as a first step. Q should be followed with other measures of heterogeneity. Options include Tau2, which is an estimate of the variance of the true effect sizes from observed data—it is expressed on the same scale as the observed effect sizes and can thus be compared relative to the overall effect size estimate. Another option is I2, which is an expression (as a percentage) of how much of the observed variation in effect sizes can be traced back to heterogeneity. Given the difference in coverage and meaning, a strong meta-analysis would report all three (Borenstein et al., 2021). Much of the PBL literature could be strengthened in this space as reporting is generally limited to a Q test for heterogeneity.

For the example derived from Leary et al. (2013), the overall heterogeneity is statistically significant, $Q(352) = 1756.83$, $p = 0.01$. It is quite large ($Tau2 = 0.13$) relative to the overall effect size ($g = 0.27$), and a large portion of that observed variability ($I2 = 80.0\%$) can be attributed to heterogeneity in observed effect sizes. In short, there is a large degree of heterogeneity in these results that warrant additional analyses, especially among the studies that attempt to account for the reasons behind these differences.

Publication Bias and Outliers

Determining if the observed studies represent the inherent bias of academia to favor statistically significant results is another important determination. A rather dated technique that is reflected in several PBL meta-analyses is vote counting positive and negative results. A vote count is not recommended as it will generally lack sensitivity (Borenstein et al., 2021). Much of that has to do with authors engaging in directional hypothesis testing, so a lack of a positive finding for PBL is often classified the same as a finding that statistically favors the control group. Instead, publication bias is best assessed with a funnel plot and Egger's test. It should

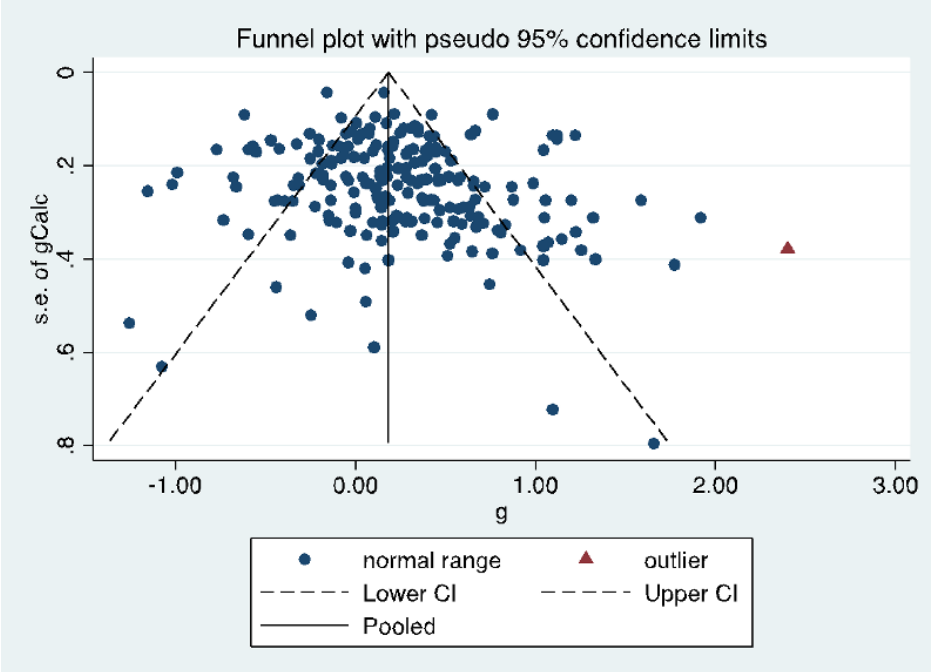


Figure 3. Sample Funnel Plot with Trimmed Outlier Based on Leary et al. (2013)

be noted that the original Leary et al. (2013) study did not address publication bias beyond a vote count in a related conference paper. The funnel plot (see Figure 3) represents a new analysis, and plots the standard error on the y-axis and effect sizes on the x-axis.

Visual inspection of the funnel plot as well as the result of an Egger's test (see Egger et al., 1997) both suggest that there is publication bias for these data, $t(352) = 5.15$, $p < .001$. To underscore Borenstein et al. (2021), a vote count based on these same data for a conference paper did not find publication bias. As recommended, a further exploration of these data was conducted and revealed a single outlier (defined as three standardized deviation units above the mean). There were no systematic explanations as to why the outlier or other high value or high standard error studies were unusual, relative to the other included studies. As a result, the outlier effect size was trimmed to the next highest value, and it was kept.

A re-analysis indicates that there was still publication bias, so one limitation is that the observable outcome data from empirical studies likely overestimates the effect of PBL. To quantify that overestimation, a trim-and-fill analysis was performed (see Figure 4). To completely nullify the publication bias, another 77 imputed studies would need to be added, dropping the overall effect size by almost two-thirds to $g = 0.103$

As an update from Leary et al (2013) analyses, which did not examine outliers, in this paper an outlier was found and trimmed. Also, while the prior vote count was not sensitive to existing publication bias, a decision was made to not numerically correct for the newly found bias but rather keep results on the same scale both to the prior study and to the data itself. Maintaining the scale allows readers to compare summary findings to the 2013 analysis or unadjusted summary findings to individual outcome effect sizes. Readers are encouraged to keep in mind the degree of overestimation for results from this paper and the likely overestimation of Leary et al. (2013), as well as the importance of using modern and multiple methods in examining it.

Robust Variance Estimation

One of the challenges with meta-analysis is the independence assumption. As shown in Figure 1, it is common for any one study to have multiple outcomes. For Leary et al. (2013) and the additional data across 113 studies, there were anywhere from 1 to 24 outcomes with an average of 3.12 outcomes per study. Researchers are left with several unpalatable choices, from combining effect sizes that may have very

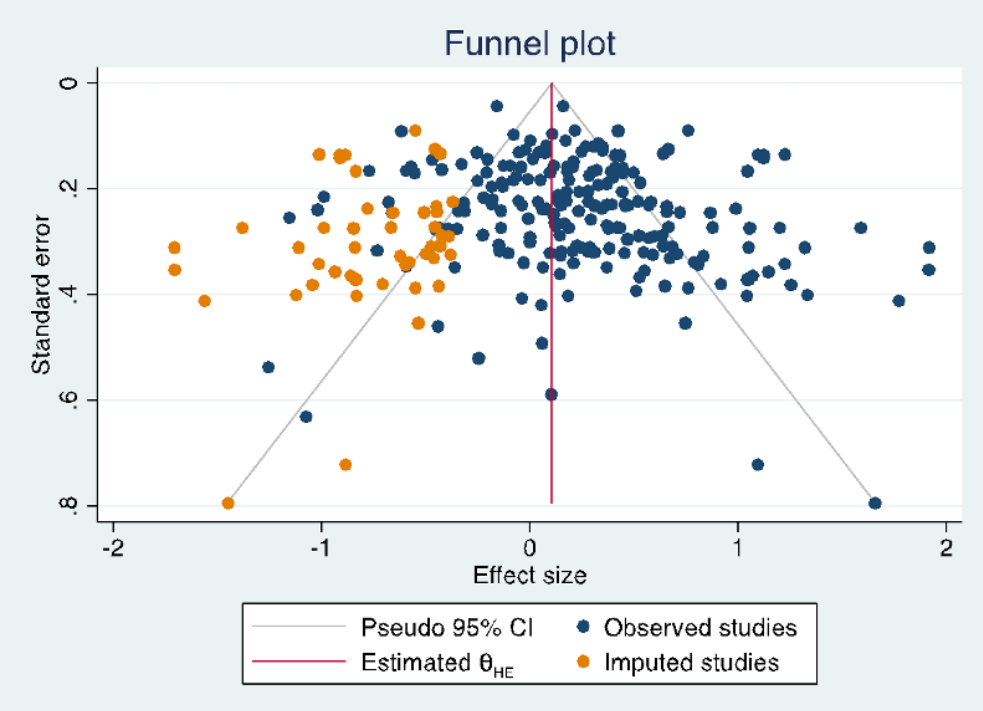


Figure 4. Sample Trim-and-Fill to Quantify Publication Bias Based on Leary et al. (2013)

different coding characteristics, to selecting the most germane effect, which then ignores data (Cooper et al., 2019), or violating the independence assumption, as most meta-analyses choose to do. It should be noted that the best technique, an HLM meta-analysis (Kalaian et al., 1999) which accounts for the nested nature of data, is appropriate only in a shared measurement context, such as a standardized test. Given the diverse contexts in which PBL has been used, another alternative is using robust variance estimate (RVE) to test the assumption that study of origin is associated with variance in outcomes (Hedges et al., 2010). In a new analysis of Leary et al. (2013) data, a range of assumed rho values were tested and used to model Hedges' g. At the two extremes of no correlation (rho = 0) and almost perfect correlation (rho = 0.99), Hedges' g was identical to four decimal places, suggesting that the study of origin does not make an impact on the final effect size. These preliminary analyses of heterogeneity, publication bias, and addressing the nested nature of the data may not directly address any research questions, but they are important and also good ways to improve on the contributions of existing PBL meta-analyses.

Forest Plot

Any assessment of the overall effect size of an intervention is best presented alongside a forest plot. To begin, authors should be clear about the modeling assumption they are making; this should be done a priori based on the nature of their included studies. A fixed effects model assumes that there is a single true effect size.

At 353 total outcomes, the forest plot for Leary et al. (2013) is so large that it is challenging to see meaningful details. Instead of showing that, a forest plot of just the

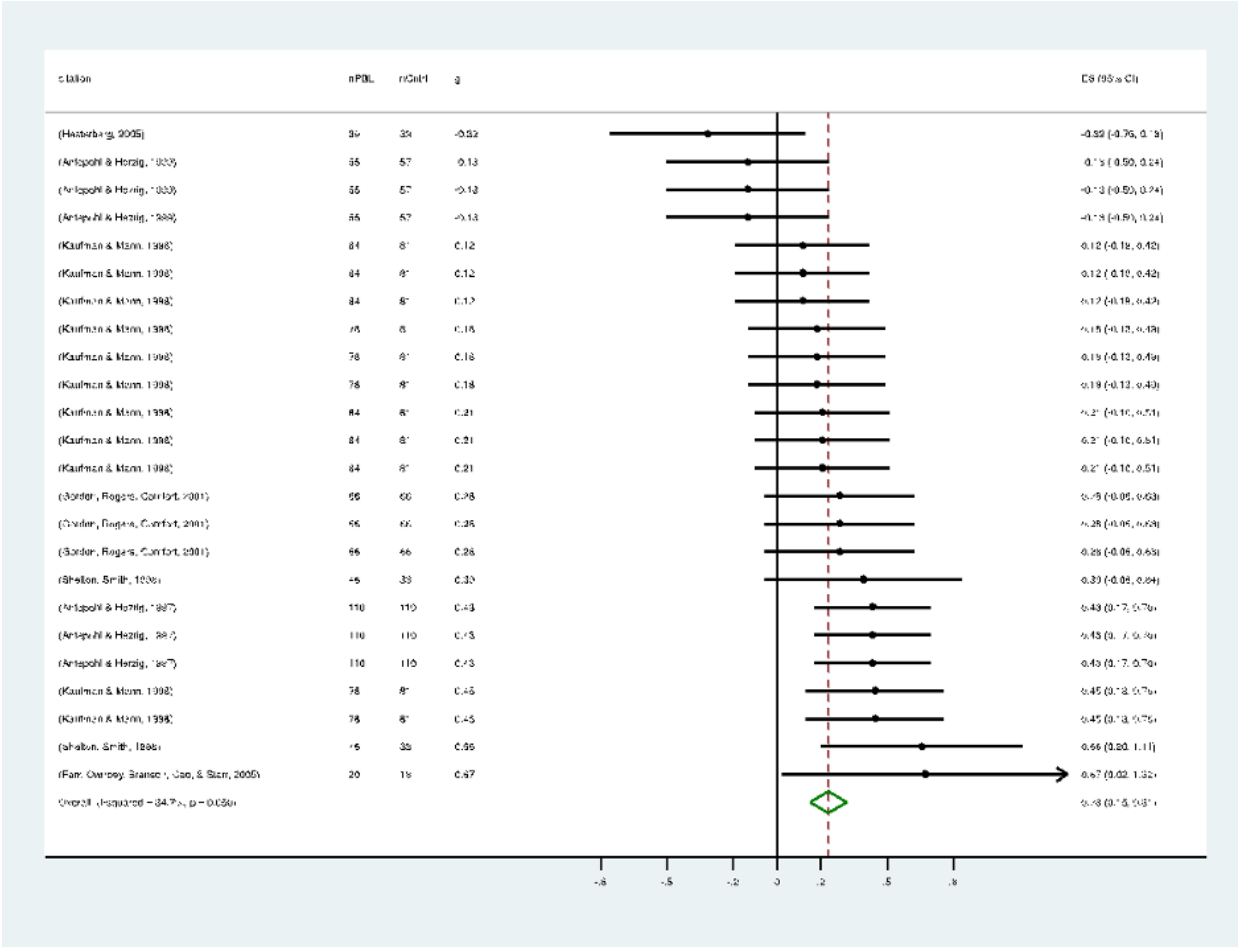


Figure 5. Forest Plot of Content Novice Tutors Only from Leary et al. (2013)

outcomes that used a content novice as the tutor is shown (see Figure 5). The overall estimate (expressed as a green diamond) shows the overall effect size (at the apex) and the confidence interval (at the width). It can be paired with a vertical red-dashed line extending from the overall effect size to visualize how much lower or higher each individual estimate is from this mean. To frame these results in terms of the research question, the overall contribution of PBL is modest (g = 0.27), even in the likely over-estimated form. There is a massive range of outcomes from LeJeune (2002), g = -1.26, to the promising results of Ceconi (2006), g = 1.91. The large variation and previously reported heterogeneity in the results and precision of these findings call for identification of systematic explanations and patterns in these studies.

Individual citations for the study of origin, the sample size of the PBL, control conditions, and both a visual as well as numerical report of the individual effect size (small dot) and upper/lower confidence interval (outer edges of the line) are shown for each outcome in Figure 5. Sorting by the effect size estimate gives a sense for the range of scores. The last outcome has a confidence interval beyond what can be shown, which is indicated by the arrow.

Future Meta-Analytic PBL Research

This paper is a necessarily terse discussion of the best meta-analysis practices in the context of existing PBL literature. There is clearly room to adopt more modern techniques in the exploration of a literature base that has only expanded since its inception in the late sixties. It is important for PBL researchers to refer to existing meta-analyses and continue to conduct them moving forward, especially as the meta-analytic methods are refined and improved. It may also be time for another meta-synthesis review, especially one that can incorporate reviews where PBL is clearly included and uniquely reported on or featured.

Currently, there are no meta-analyses in PBL focused on correlational designs or using Bayesian network meta-analysis. A review of standardized assessments in PBL in various disciplines would be intriguing. It could focus on targeted groups of studies or common assessments. Alternatively, a common vocabulary for discussing PBL interventions could be developed, perhaps as a revisioning of Barrow's taxonomy (1986). This undertaking could be robust and a wonderful addition to the PBL literature if a meta-analytic method is used. Finally, no authors have undertaken a registered Campbell Collaboration systematic review of PBL literature, which could help raise the profile of scholarship in this space (see Campbell Systematic Reviews, https://onlinelibrary.wiley.com/journal/18911803).

In PBL, much work remains. Meta-analysis holds a key to understanding the needs and gaps in PBL activities, as well as how to improve the practice, by analyzing multiple studies. As an established method for comparing empirical research across disciplines, meta-analysis can contribute greatly to understanding of PBL and the best directions for future research and practice.

References

Albanese, M. (2000). Problem-based learning: Why curricula are likely to show little effect on knowledge and clinical skills. Medical Education, 34(9), 729–738.

Albanese, M., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. Academic Medicine, 68(1), 52–81.

Barrows, H. S. (1986). A taxonomy of problem-based learning methods. Medical Education, 20(6), 481–486.

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. New Directions for Teaching and Learning, 68, 3–12.

Belland, B., Walker, A., & Kim, N. (2017a). A Bayesian network meta-analysis to synthesize the influence of contexts of scaffolding use on cognitive outcomes in STEM education. Review of Educational Research, 87(6), 1042–1081. https://doi.org/10.3102/0034654317723009

Belland, B., Walker, A., Kim, N., & Lefler, M. (2017b). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. Review of Educational Research, 82(2), 309–344. https://doi.org/10.3102/0034654316670999

Berkson, L. (1993). Problem-based learning: Have the expectations been met? Academic Medicine, 68(10, Suppl.), S79-88.

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. Educational Researcher, 34(6), 3–15. https://doi.org/10.3102/0013189X034006003

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). Introduction to meta-analysis (2nd ed.). Wiley.

Ceconi, A. (2006). Influence of problem-based learning instruction on decision-making skills in respiratory therapy students [Dissertation]. Seton Hall University.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence

Colliver, J. A. (2000). Effectiveness of problem-based learning curricula: Research and theory. Academic Medicine, 75(3), 259–266. Earlbaum Associates.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). The handbook of research synthesis and meta-analysis

(3rd ed.). Russell Sage Foundation.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. Learning and Instruction, 13(5), 533–568. https://doi.org/10.1016/S0959-4752(02)00025-7

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ, 315(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. Review of Educational Research, 75, 27–61. https://doi.org/10.3102/00346543075001027

Glanville, J. (2019). Searching bibliographic databases. In H. Cooper, L. Hedges, & J. Valentine (Eds.), The handbook of research synthesis and meta-analysis (3rd ed., pp. 73–100). Russell Sage Foundation.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5(10), 3–8.

Glass, G. V. (2006). Meta-analysis: The quantitative synthesis of research findings. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), Handbook of complementary methods in education research (pp. 427–438). Lawrence Erlbaum Associates.

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Research Synthesis Methods, 11(2), 181–217. https://doi.org/10.1002/jrsm.1378

Hansen, C., Steinmetz, H., & Block, J. (2022). How to conduct a meta-analysis in eight steps: A practical guide. Management Review Quarterly, 72(1), 1–19. https://doi-org.dist.lib.usu.edu/10.1007/s11301-021-00247-4

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods, 1, 39–65. https://doi.org/10.1002/jrsm.5

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). Educational Psychologist, 42(2), 99–107. https://doi.org/10.1080/00461520701263368

Jonassen, D. H. (2000). Toward a design theory of problem solving. Educational Technology Research and Development, 48(4), 63–85. https://doi-org.dist.lib.usu.edu/10.1007/BF02300500

Kalaian, H. A., Mullan, P. B., & Kasim, R. M. (1999). What can studies of problem-based learning tell us? Synthesizing and modeling PBL effects on National Board of medical examination performance: Hierarchical linear modeling meta-analytic approach. Advances in Health Sciences Education, 4(3), 209–221. https://doi-org.dist.lib.usu.edu/10.1023/A:1009871001258

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. Educational Researcher, 49(4), 241–253. https://doi.org/10.3102/0013189X20912798

Leary, H., & Walker, A. (2018). Meta-analysis and meta-synthesis methodologies: Rigorously piecing together research. TechTrends, 62(5), 525–534. https://doi.org/10.1007/s11528-018-0312-7

Leary, H., Walker, A., Shelton, B., & Fitt, M. (2013). Exploring the relationships between tutor background, tutor training, and student learning: A problem-based learning meta-analysis. Interdisciplinary Journal of Problem-Based Learning, 7(1). https://doi.org/10.7771/1541-5015.1331

LeJeune, N. (2002). Problem-based learning instruction versus traditional instruction on self-directed learning, motivation, and grades of undergraduate computer science students [Doctoral dissertation]. University of Colorado at Denver.

Lipsey, M., & Wilson, D. (2001). Practical meta-analysis. Sage.

Newman, M. (2003). A pilot systematic review and meta-Analysis on the effectiveness of problem based learning. The Campbell Collaboration.

Rosenthal, R. (1979). The 'file drawer problem' and tolerance for null results. Psychological Bulletin, 86, 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Strobel, J., & Barneveld, A. V. (2009). Is PBL effective? A meta-synthesis of meta-analyses comparing problem-based learning to conventional classroom learning. Interdisciplinary Journal of Problem-Based Learning, 3(1), 44–58. https://doi.org/10.7771/1541-5015.1046

Sugrue, B. (1993). Specifications for the design of problem-solving assessments in science. National Center for Research on Evaluation, Standards, and Student Testing.

Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem solving ability. Educational Measurement: Issues and Practice, 14(3), 29–36. https://doi.org/10.1111/j.1745-3992.1995.tb00865.x

Vernon, D. T., & Blake, R. L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. Academic Medicine, 68(7), 550–563. https://doi.org/10.1097/00001888-199307000-00015

Wilson, D. (n.d.). Practical Meta-Analysis Effect Size Calculator. Retrieved October 27, 2023, from https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-R-main.php.

Walker, A., & Leary, H. (2009). A problem-based learning meta analysis: Differences across problem types, implementation types, disciplines, and assessment levels.

Interdisciplinary Journal of Problem-Based Learning, 3(1), 12–43. https://doi.org/10.7771/1541-5015.1331

Author Bios

Andrew Walker is faculty in and Department Head of Instructional Technology and Learning Sciences at Utah State University. His research has a common theme of using problem-centered pedagogies like problem-based learning to create and examine learning experiences. That work often expands to closely related interventions such as scaffolding and educational games; He explores these interests through various methods such as meta-analysis techniques including traditional, network, and Bayesian network meta-analysis as well as multilevel modeling.

Heather Leary is an Associate Professor of Instructional Psychology & Technology at Brigham Young University. Her research focuses on tackling complex educational problems of practice in K-12 and higher education. She works collaboratively with practitioners in various disciplines (e.g., science, math, arts) and spaces (e.g., online environments, professional development) to iteratively design and implement potential solutions. She uses design-based research, research-practice partnerships, problem-based learning, and research synthesis to generate usable knowledge and alignment between the principles of learning and instructional practice.

Acknowledgements

Thanks to Bonnie Jones for demonstrating the use of rayyan for scoping reviews.